US009215181B2

(12) **United States Patent**
Toy

(10) **Patent No.: US 9,215,181 B2**
(45) **Date of Patent: Dec. 15, 2015**

(54) **SYSTEMS AND METHODS FOR MANAGING A NETWORK**

(71) Applicant: **Comcast Cable Communications, LLC**, Philadelphia, PA (US)

(72) Inventor: **Mehmet Toy**, Allendale, NJ (US)

(73) Assignee: **Comcast Cable Communications, LLC**, Philadelphia, PA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 125 days.

(21) Appl. No.: **13/669,742**

(22) Filed: **Nov. 6, 2012**

(65) **Prior Publication Data**

US 2014/0126368 A1 May 8, 2014

(51) **Int. Cl.**

H04L 12/26	(2006.01)
H04J 3/14	(2006.01)
H04J 3/24	(2006.01)
H04L 12/56	(2006.01)
H04L 12/825	(2013.01)

(52) **U.S. Cl.**

CPC ***H04L 47/26*** (2013.01); ***H04L 43/0882*** (2013.01)

(58) **Field of Classification Search**

CPC ... H04L 12/26; H04L 12/1256; H04L 47/122; H04L 12/56; H04L 12/28; H04J 3/14; H04J 1/16; H04J 3/24

USPC 370/229–235, 389, 466, 236, 246

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,782,906	B2 *	8/2010	Zhai	370/474
7,961,621	B2 *	6/2011	Bergamasco et al.	370/236
8,279,875	B2 *	10/2012	Zhu et al.	370/395.5
8,660,075	B2 *	2/2014	Ghanwani et al.	370/329
2006/0203730	A1 *	9/2006	Zur	370/235
2011/0051723	A1 *	3/2011	Rabie et al.	370/389
2013/0185454	A1 *	7/2013	Guillemin et al.	709/235

OTHER PUBLICATIONS

Ramakrishnan et al., IETF RFC 3178, Sep. 2001—The Addition of Explicit Congestion Notification (ECN) to IP.*

IEEE 802.1QauTM—2010 IEEE Standard for Local and Metropolitan area network—Virtual Local Area Networks, Amendment 13: Congestion Notification, Apr. 23, 2010, IEEE Computer Society, LAN/MAN Standards Committee.*

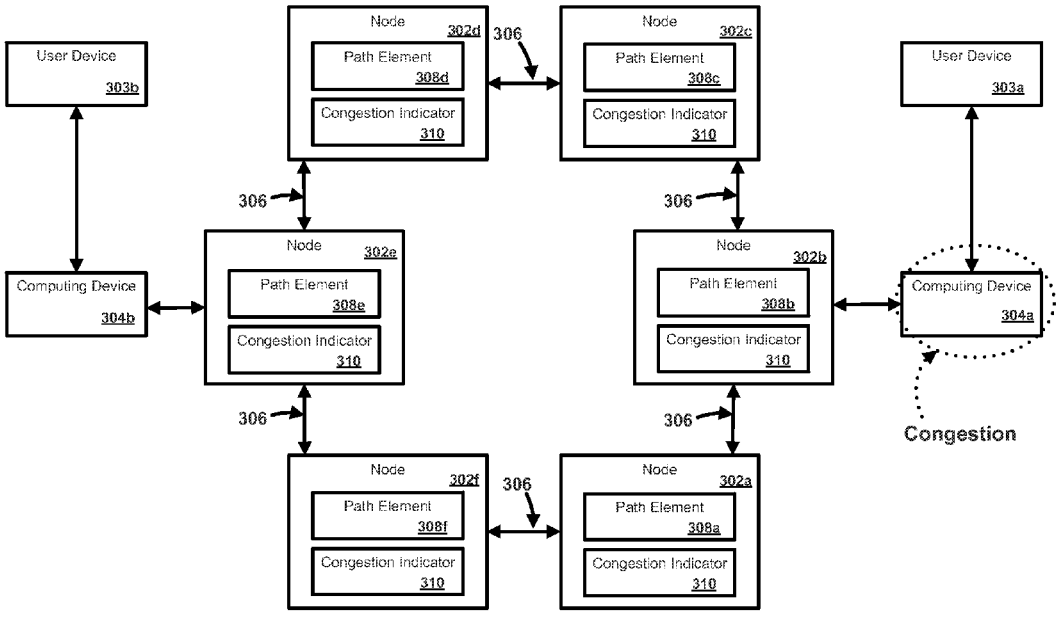
* cited by examiner

Primary Examiner — Chi H Pham
Assistant Examiner — Shailendra Kumar
(74) *Attorney, Agent, or Firm* — Ballard Spahr LLP

(57) **ABSTRACT**

Systems and methods for managing congestion in a network are disclosed. One method can comprise receiving a service tag at a first node, the service tag representing congestion information of at least a portion of the network. If the first node is a boundary node, the method comprises modifying a downstream data rate based upon the congestion information, and if the first node is not a boundary node, the method comprises transmitting the congestion information to a second node.

32 Claims, 8 Drawing Sheets



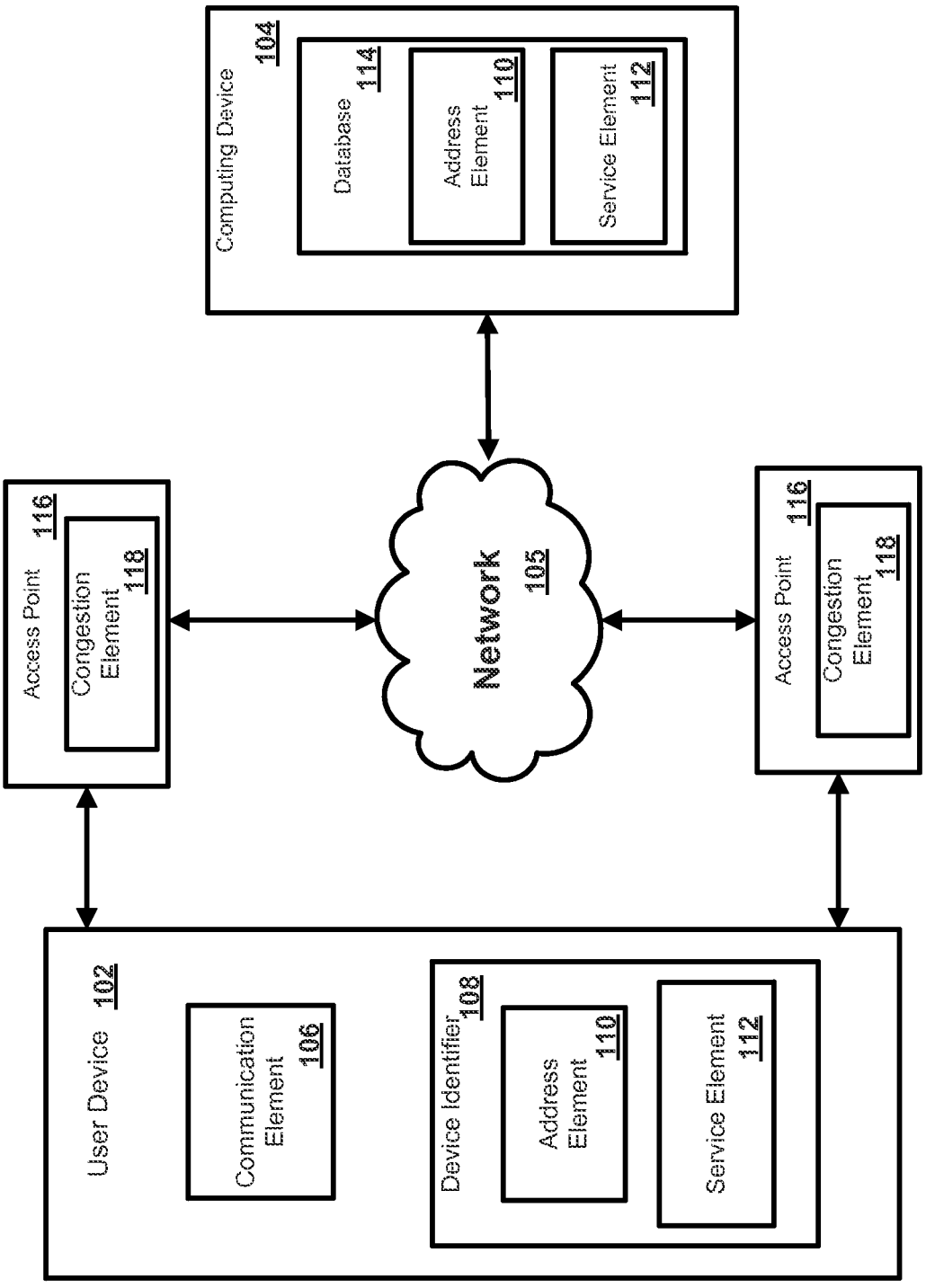

Computing Device 104
Database 114
Address Element 110
Service Element 112
Network 105
Access Point 116
Congestion Element 118
Access Point 116
Congestion Element 118
User Device 102
Communication Element 106
Device Identifier 108
Address Element 110
Service Element 112

FIG. 1A

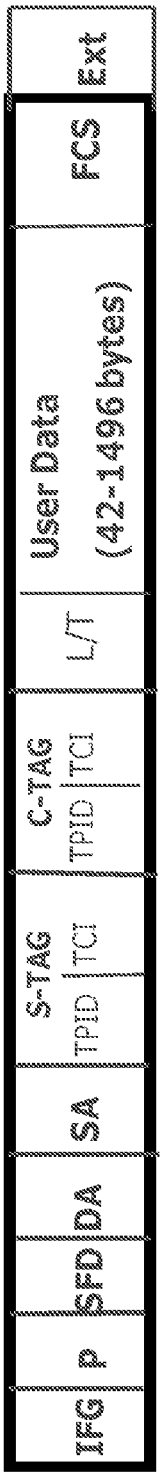


FIG. 1B

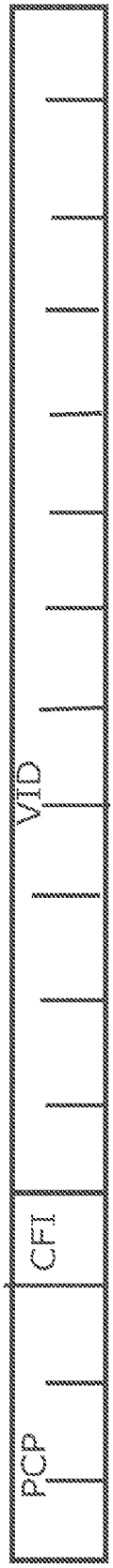


FIG. 1C

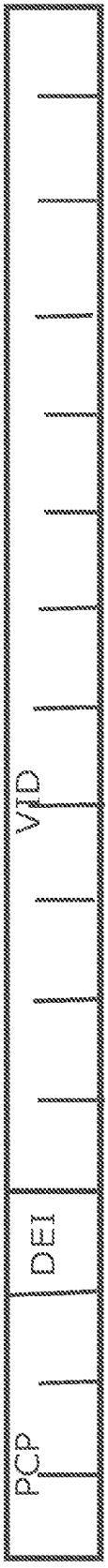


FIG. 1D

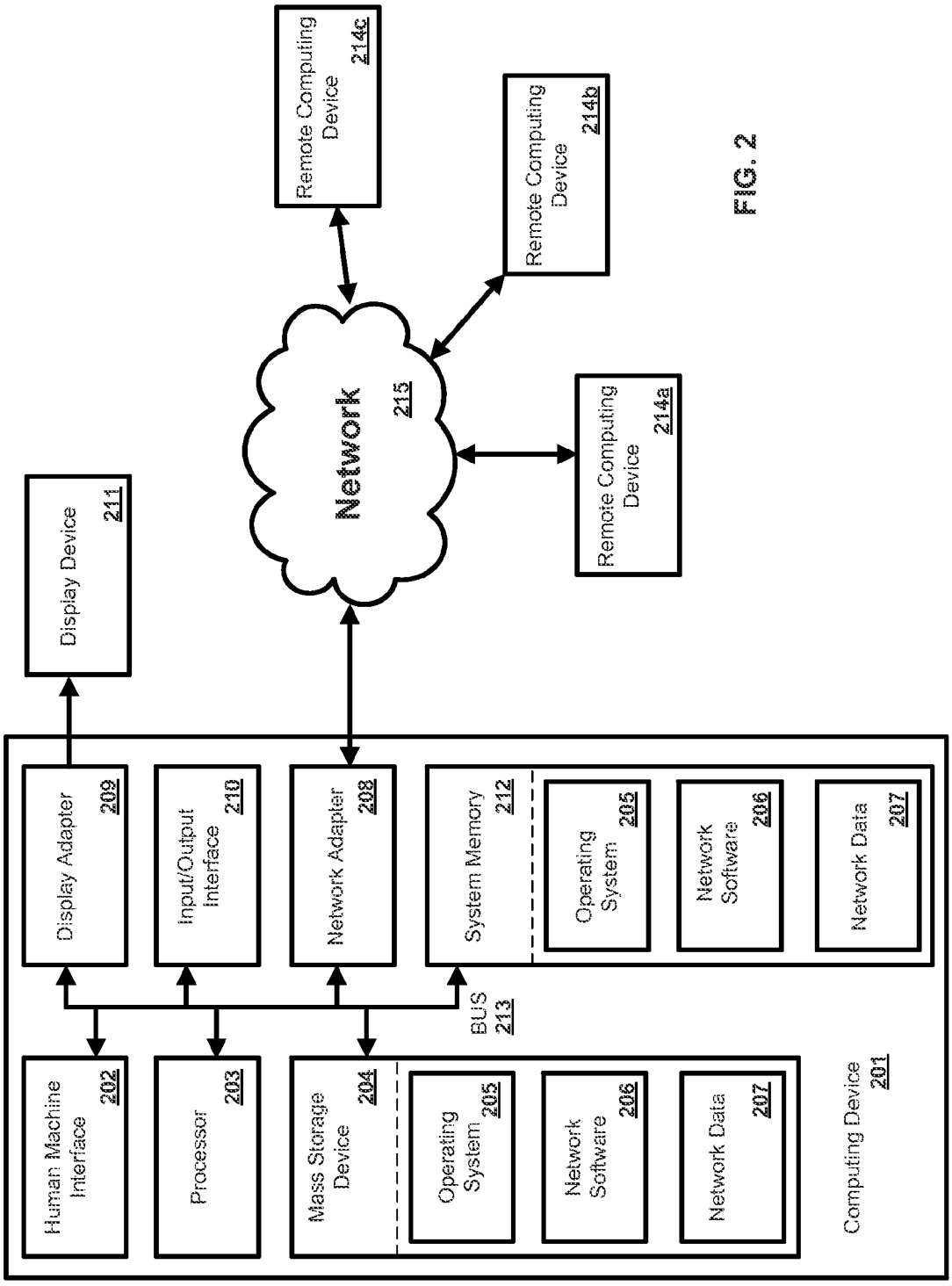

Remote Computing Device 214c
Remote Computing Device 214b
Remote Computing Device 214a
Network 215
Display Device 211
Display Adapter 209
Input/Output Interface 210
Network Adapter 208
System Memory 212
Operating System 205
Network Software 206
Network Data 207
BUS 213
Human Machine Interface 202
Processor 203
Mass Storage Device 204
Operating System 205
Network Software 206
Network Data 207
Computing Device 201

FIG. 2

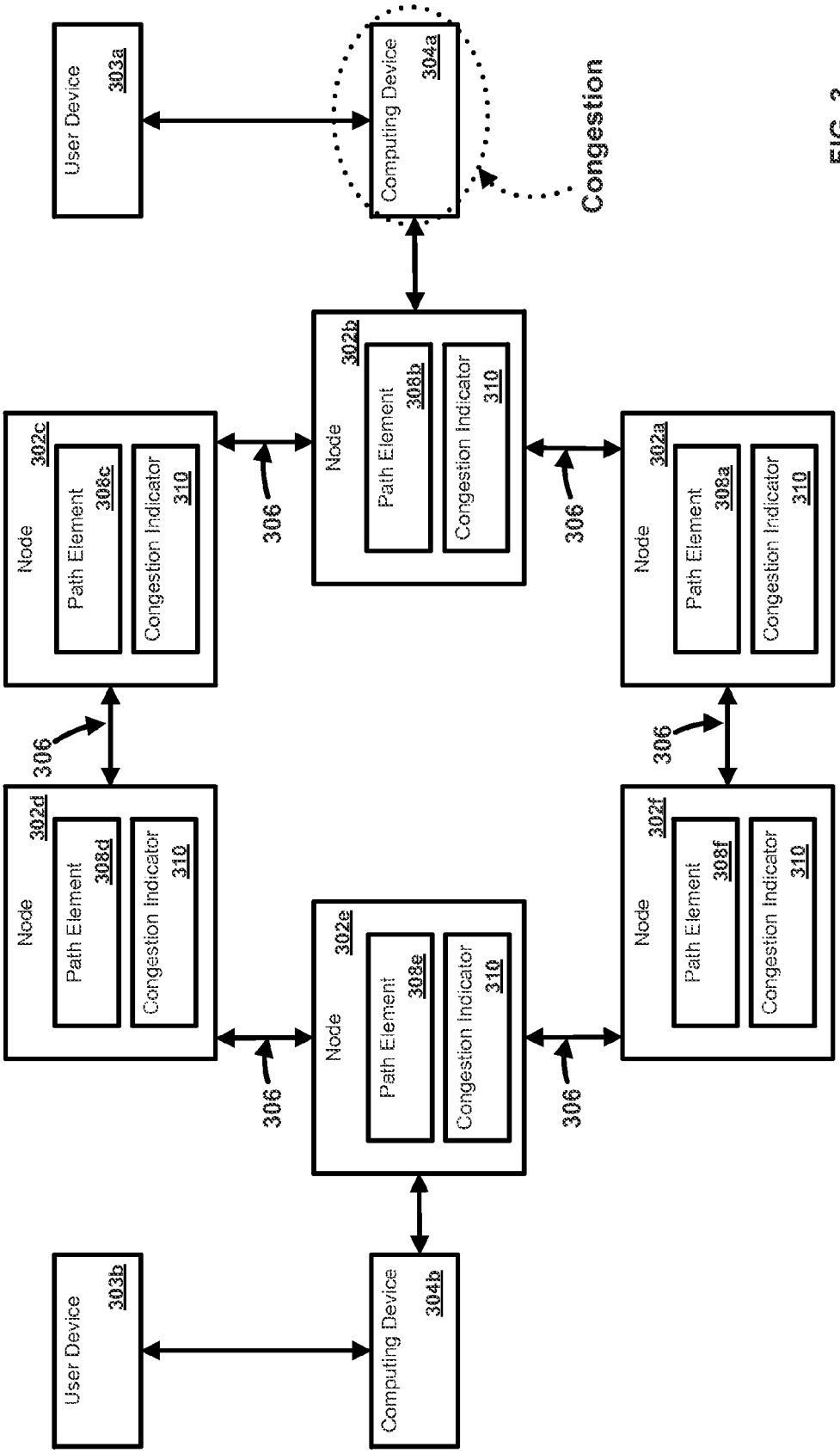

User Device 303a
Computing Device 304a
Congestion
Node 302b
Path Element 308b
Congestion Indicator 310
Node 302c
Path Element 308c
Congestion Indicator 310
Node 302a
Path Element 308a
Congestion Indicator 310
Node 302d
Path Element 308d
Congestion Indicator 310
Node 302e
Path Element 308e
Congestion Indicator 310
Node 302f
Path Element 308f
Congestion Indicator 310
306
User Device 303b
Computing Device 304b

FIG. 3

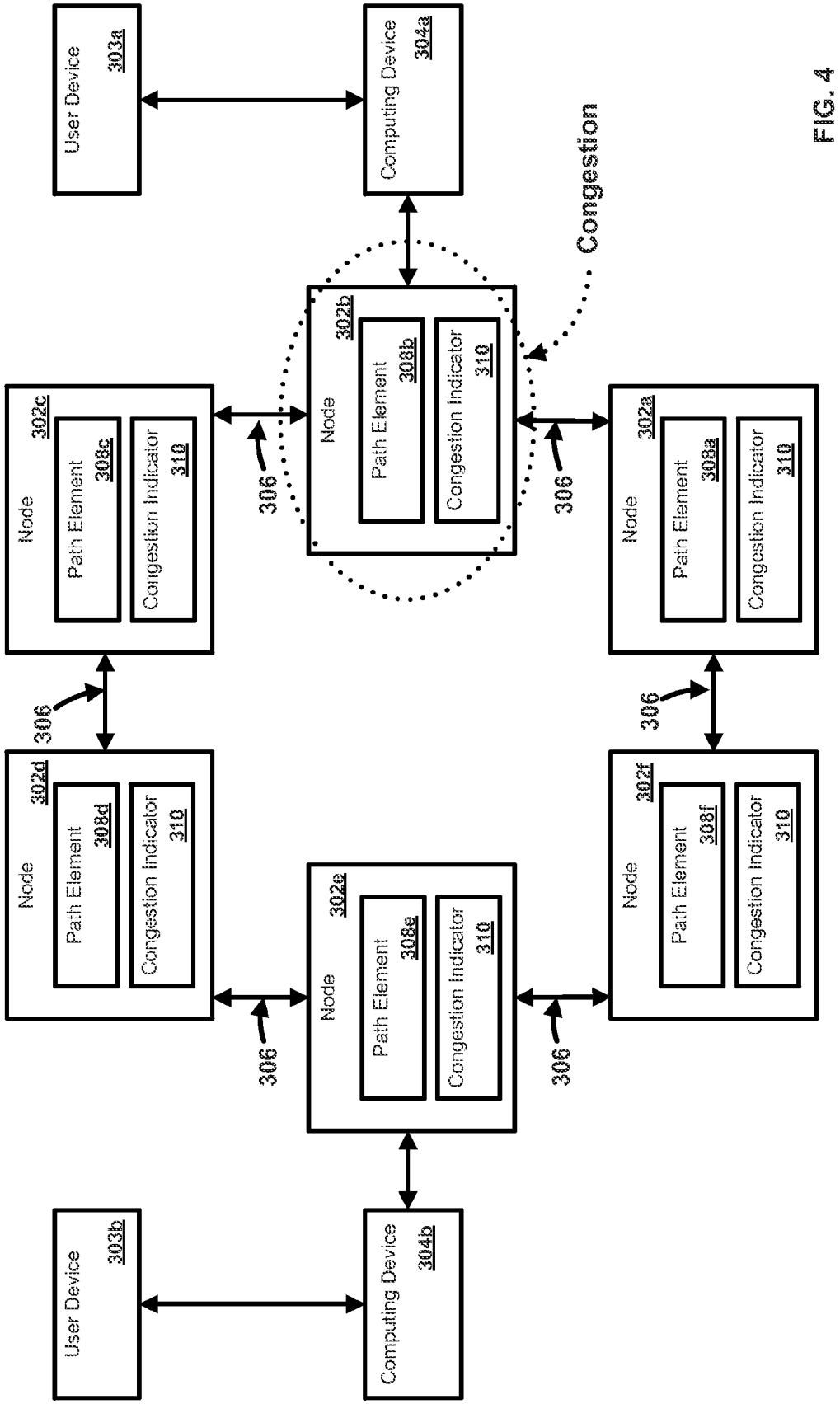

User Device 303a
Computing Device 304a
Congestion
Node 302b
Path Element 308b
Congestion Indicator 310
Node 302c
Path Element 308c
Congestion Indicator 310
Node 302a
Path Element 308a
Congestion Indicator 310
Node 302d
Path Element 308d
Congestion Indicator 310
Node 302e
Path Element 308e
Congestion Indicator 310
Node 302f
Path Element 308f
Congestion Indicator 310
306
User Device 303b
Computing Device 304b

FIG. 4

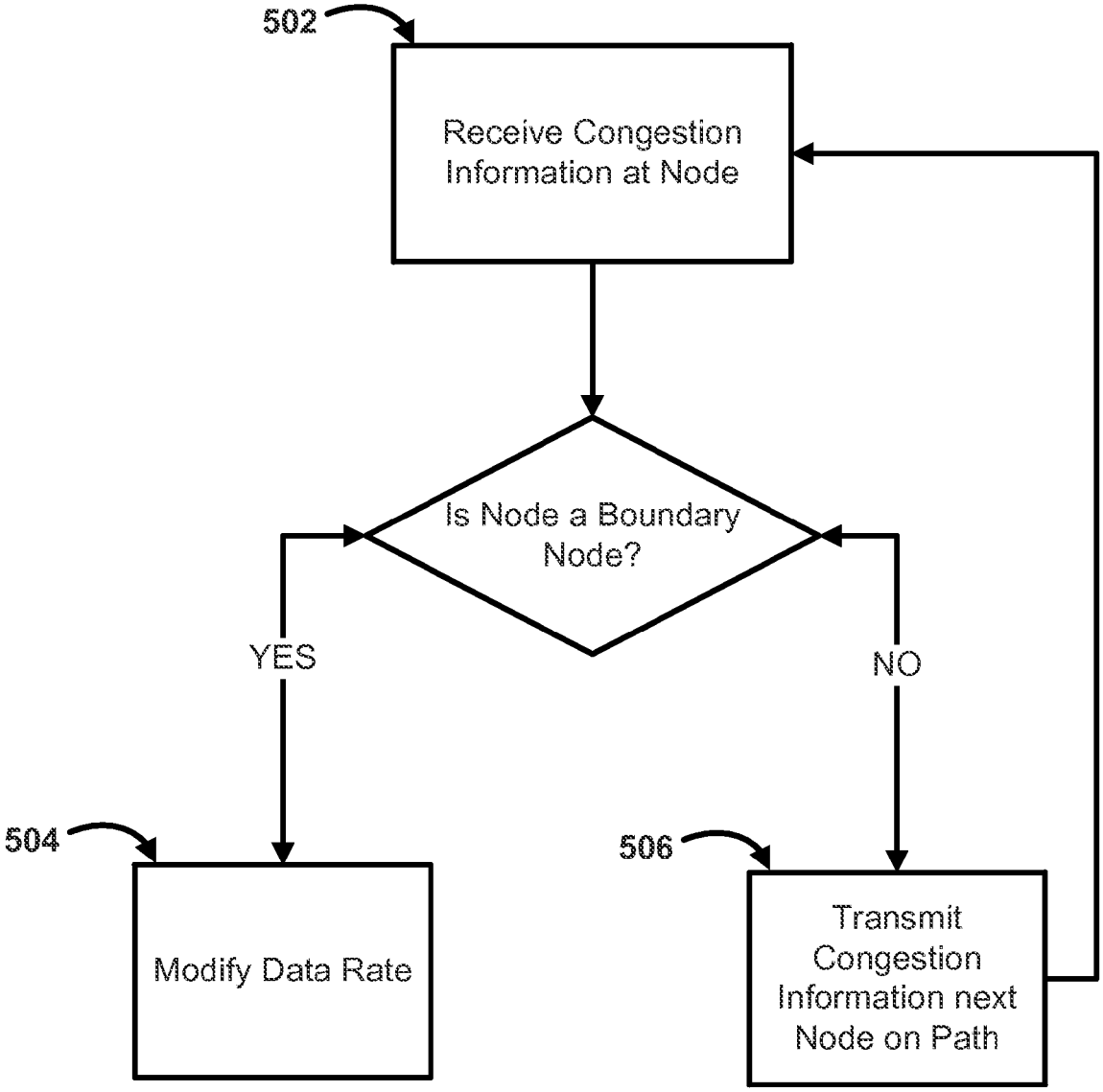

502
Receive Congestion Information at Node
Is Node a Boundary Node?
YES
NO
504
Modify Data Rate
506
Transmit Congestion Information next Node on Path

FIG. 5

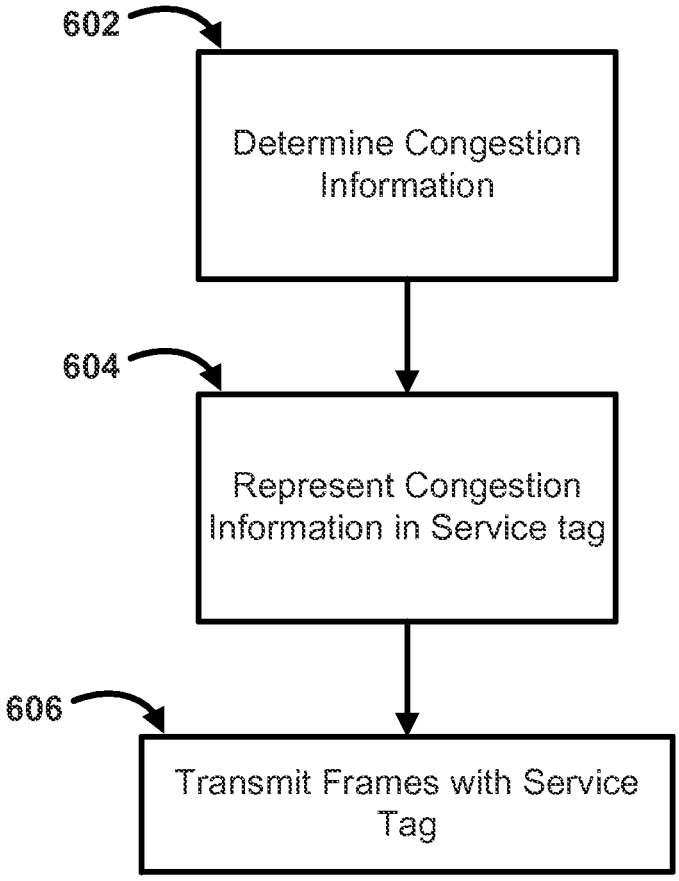

602
Determine Congestion Information
604
Represent Congestion Information in Service tag
606
Transmit Frames with Service Tag

FIG. 6

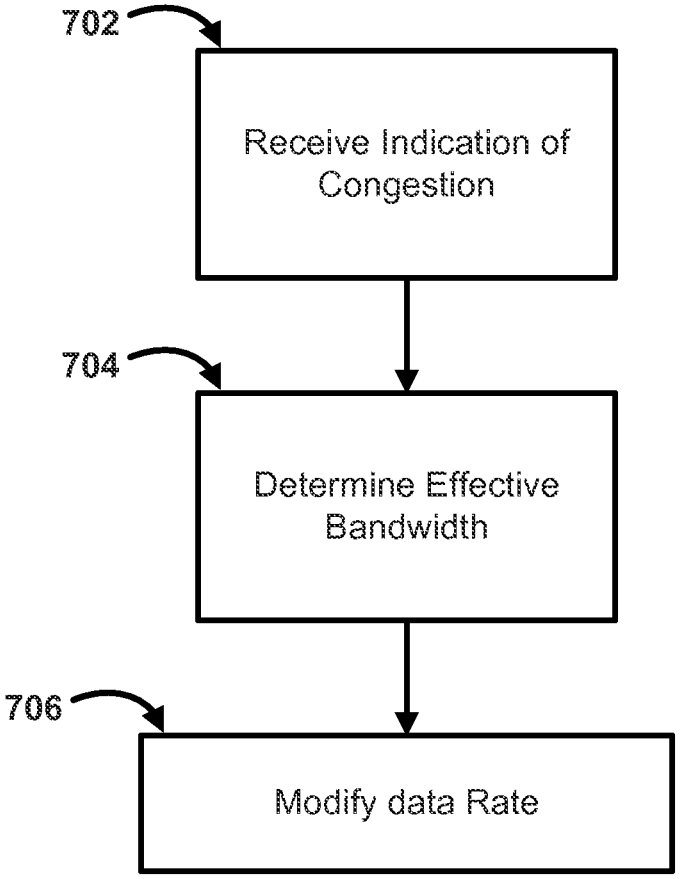

702
Receive Indication of Congestion
704
Determine Effective Bandwidth
706
Modify data Rate

FIG. 7

SYSTEMS AND METHODS FOR MANAGING A NETWORK

BACKGROUND

Networks can experience problems due to network constraints such as congestion. Certain network systems can monitor network conditions, but such systems suffer from deficiencies. Deficiencies can include failing to provide end-to-end and/or domain congestion indication in a granular fashion such as per port, per connection, or per class of service. This disclosure addresses such and other shortcomings related to network management.

SUMMARY

It is to be understood that both the following general description and the following detailed description are exemplary and explanatory only and are not restrictive, as claimed. Methods and systems for managing a network are disclosed. The methods and systems described herein, in one aspect, can manage congestion in a network.

In an aspect, discard eligibility (DE) bits in a service tag (S-Tag) of one or more data packets can be used to indicate congestion end-to-end or on a domain basis in packet networks. Data packets can comprise Ethernet frames. For example, when a downstream node is congested, the downstream node can define a congestion indicator (e.g., a. DE bit) in one or more frames. The one or more frames comprising the congestion indicator can be transmitted to one or more upstream nodes to indicate congestion. As a further example, the congestion indicator can be a bit value such as the DE bit set to a value of "1." In an aspect, the congestion indicator can be used to indicate congestion information per port, per connection, or per class of service, or a combination thereof.

In an aspect, methods can comprise receiving a frame with a service tag at a first node. The service tag can represent congestion information of at least a portion of the network. If the first node is a boundary node, a downstream data rate can be modified based upon the congestion information. If the first node is not a boundary node, the congestion information can be transmitted to a second node. A boundary node can be a node that initiates or terminates congestion control. For example, boundary nodes can support Ethernet virtual connection termination points, such as, customer premises equipment. As a further example, for a domain defined by edge routers, boundary nodes can comprise the provider edge (PE) routers.

In another aspect, the methods can comprise determining congestion information of at least a portion of a network. User frames with a service tag can be transmitted to an upstream device of the network to indicate congestion.

In a further aspect, the methods can comprise receiving and addressing an indication of network congestion. The indication can relate to, for example, a service flow. A service flow can comprise an end-to-end traffic flow (e.g., from customer premises equipment (CPE) to other CPE or network devices) defined by traffic parameters such as average output rate, maximum output burst, and the like. As an example, a service flow can comprise an Ethernet virtual connection between user network interfaces (UNI) of a network. As a further example, a service flow can comprise a group of packets/frames flowing in an Ethernet virtual connection between UNIs and belong to an application with a defined class of service. An aggregate service flow can comprise one or more service flows.

As such, an effective bandwidth can be determined for the service flow. Further, a downstream data rate associated with the service flow can be modified based upon the effective bandwidth determined.

Additional advantages will be set forth in part in the description which follows or may be learned by practice. The advantages will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive, as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments and together with the description, serve to explain the principles of the methods and systems:

FIG. **1**A is a block diagram of an exemplary system and network;

FIG. **1**B is a block diagram of an exemplary data packet that can be transported by the network;

FIG. **1**C is a block diagram of an exemplary customer tag;

FIG. **1**D is a block diagram of an exemplary service tag;

FIG. **2** is a block diagram of an exemplary computing device;

FIG. **3** is a diagram of an exemplary system and network;

FIG. **4** is a diagram of an exemplary system and network;

FIG. **5** is a flow chart of an exemplary method;

FIG. **6** is a flow chart of an exemplary method; and

FIG. **7** is a flow chart of an exemplary method.

DETAILED DESCRIPTION

Before the present methods and systems are disclosed and described, it is to be understood that the methods and systems are not limited to specific methods, specific components, or to particular implementations. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

As used in the specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Ranges may be expressed herein as from "about" one particular value, and/or to "about" another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint.

"Optional" or "optionally" means that the subsequently described event or circumstance may or may not occur, and that the description includes instances where said event or circumstance occurs and instances where it does not.

Throughout the description and claims of this specification, the word "comprise" and variations of the word, such as "comprising" and "comprises," means "including but not limited to," and is not intended to exclude, for example, other components, integers or steps. "Exemplary" means "an example of" and is not intended to convey an indication of a preferred or ideal embodiment, "Such as" is not used in a restrictive sense, but for explanatory purposes.

Disclosed are components that can be used to perform the disclosed methods and systems. These and other components are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these components are disclosed that while specific reference of various individual and collective combinations and permutation of these may not be explicitly disclosed, each is specifically contemplated and described herein, for all methods and systems. This applies to all aspects of this application including, but not limited to, steps in disclosed methods. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific embodiment or combination of embodiments of the disclosed methods.

The present methods and systems may be understood more readily by reference to the following detailed description of preferred embodiments and the examples included therein and to the Figures and their previous and following description.

As will be appreciated by one skilled in the aft, the methods and systems may take the form of an entirely hardware embodiment, an entirely software embodiment, or an embodiment combining software and hardware aspects. Furthermore, the methods and systems may take the form of a computer program product on a computer-readable storage medium having computer-readable program instructions (e.g., computer software) embodied in the storage medium. More particularly, the present methods and systems may take the form of web-implemented computer software. Any suitable computer-readable storage medium may be utilized including hard disks. CD-ROMs, optical storage devices, or magnetic storage devices.

Embodiments of the methods and systems are described below with reference to block diagrams and flowchart illustrations of methods, systems, apparatuses and computer program products. It will be understood that each block of the block diagrams and flowchart illustrations, and combinations of blocks in the block diagrams and flowchart illustrations, respectively, can be implemented by computer program instructions. These computer program instructions may be loaded onto a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions which execute on the computer or other programmable data processing apparatus create a means for implementing the functions specified in the flowchart block or blocks.

These computer program instructions may also be stored in a computer-readable memory that can direct a computer or other programmable data processing apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory produce an article of manufacture including computer-readable instructions for implementing the function specified in the flowchart block or blocks. The computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer-implemented process such that the instructions that execute on the computer or other programmable apparatus provide steps for implementing the functions specified in the flowchart block or blocks.

Accordingly, blocks of the block diagrams and flowchart illustrations support combinations of means for performing the specified functions, combinations of steps for performing the specified functions and program instruction means for performing the specified functions. It will also be understood that each block of the block diagrams and flowchart illustrations, and combinations of blocks in the block diagrams and flowchart illustrations, can be implemented by special purpose hardware-based computer systems that perform the specified functions or steps, or combinations of special purpose hardware and computer instructions.

FIG. **1**A illustrates various aspects of an exemplary network in which the present methods and systems can operate. The present disclosure is relevant to systems and methods for managing a network, for example. Those skilled in the art will appreciate that present methods may be used in various types of networks and systems that employ both digital and analog equipment. The system is described as comprised of elements. An element can be software, hardware, or a combination of software and hardware. One skilled the art will appreciate that provided herein is a functional description and that the respective functions can be performed by software, hardware, or a combination of software and hardware.

The system and network can comprise a user device **102** in communication with a computing device **104** such as a server or Network Interface Device (NID), for example. The computing device **104** can be disposed locally, or remotely, relative to the user device **102**. As an example, the user device **102** and the computing device **104** can be in communication via a private and/or public network **105** such as the Internet. Other forms of communications can be used such as wired and wireless telecommunication channels, for example.

In an aspect, the user device **102** can be an electronic device such as a computer, a smartphone, a laptop, a tablet, a set top box, a display device, or other device capable of communicating with the computing device **104**. As an example, the user device **102** can comprise a communication element **106** for providing an interface to a user to interact with the user device **102** and/or the computing device **104**. The communication element **106** can be any interface for presenting information to the user and receiving user feedback such as a web browser (e.g., Internet Explorer, Mozilla Firefox, Google Chrome, Safari, or the like). Other software, hardware, and/or interfaces can be used to provide communication between the user and one or more of the user device **102** and the computing device **104**. As an example, the communication element **106** can request or query various files from a local source and/or a remote source. As a further example, the communication element **106** can transmit data to a local or remote device such as the computing device **104**.

In an aspect, the user device **102** can be associated with a user identifier or device identifier **108**. As an example, the device identifier **108** can be any identifier, token, character, string, or the like, for differentiating one user or user device (e.g., user device **102**) from another user or user device. In a further aspect, the device identifier **108** can identify a user or user device as belonging to a particular class of users or user devices. As a further example, the device identifier **108** can comprise information relating to the user device such as a manufacturer, a model or type of device, a service provider associated with the user device **102**, a state of the user device **102**, a locator, and/or a label or classifier. Other information can be represented by the device identifier **108**.

In an aspect, the device identifier **108** can comprise an address element **110** and a service element **112**. In an aspect, the address element **110** can be an internet protocol address, a network address, an Internet address, or the like. As an example, the address element **110** can be relied upon to establish a communication session between the user device **102** and the computing device **104** or other devices and/or networks. As a further example, the address element **110** can be used as an identifier or locator of the user device **102**.

In an aspect, the service element **112** can comprise an identification of a service provider associated with the user device **102** and/or with the class of user device **102**. As an example, the service element **112** can comprise information relating to, or provided by, a communication service provider that is providing or enabling communication services to the user device **102**. Services can be data services such as internet access, financial data transfers, or various file transfer, voice, and/or video services, or a combination thereof. As a further example, the service element **112** can comprise information relating to a preferred service provider for one or more particular services relating to the user device **102**. In an aspect, the address element **110** can be used to identify or retrieve the service element **11**, or vise versa. As a further example, one or more of the address element **110** and the service element **112** can be stored remotely from the user device **102** and retrieved by one or more devices such as the user device **102** and the computing device **104**. Other information can be represented by the service element **112**.

In an aspect, the computing device **104** can be a server for communicating with the user device **102**. As an example, the computing device **104** can communicate with the user device **102** for providing services. In an aspect, the computing device **104** can allow the user device **102** to interact with remote resources such as data, devices, and files. As an example, the computing device can be configured as central location, a headend, or processing facility, which can receive content (e.g., data, input programming) from multiple sources. The computing device **104** can combine the content from the various sources and can distribute the content to user locations via a distribution system.

In an aspect, the computing device **104** can manage the communication between the user device **102** and a database **114** for sending and receiving data therebetween. As an example, the database **114** can store a plurality of files, webpages, user identifiers or records, or other information. As a further example, the user device **102** can request and/or retrieve a file from the database **114**. In an aspect, the database **114** can store information relating to the user device **102** such as the address element **110** and/or the service element **112**. As an example, the computing device **104** can obtain the device identifier **108** from the user device **102** and retrieve information from the database **114** such as the address element **110** and/or the service elements **112**. As a further example, the computing device **104** can obtain the address element **110** from the user device **102** and can retrieve the service element **112** from the database **114**, or vice versa. Any information can be stored in and retrieved from the database **114**. The database **114** can be disposed remotely from the computing device **104** and accessed via direct or indirect connection. The database **114** can be integrated with the computing system **104** or some other device or system.

In an aspect, one or more access points **116** can be in communication with network **105**. As an example, one or more of the access points **116** can facilitate the connection of a device, such as user device **102**, to the network **105**. As a further example, one or more of the access points **116** can be configured as a virtual local area network (VLAN) access point. In an aspect, one or more access points **116** can be configured as part of a Carrier Ethernet Network.

In an aspect, one or more of the access points **116** can comprise a congestion element **118**. As an example, the congestion element **118** can be configured to receive/transmit data in packets or Ethernet frames. As a further example, the congestion element **118** can be configured to analyze the frames to determine downstream congestion information or transmit frames comprising a congestion indicator. In an aspect, the congestion element **118** can comprise a traffic conditioning element or function facilitates configured to analyze and condition data packets.

In an aspect, DE bits in a service-Tag (S-Tag) of Ethernet frames can be used to indicate congestion end-to-end or on a domain basis. For example, when a downstream node is congested, the node can set define a congestion indicator in one or more frames. The one or more frames comprising the congestion indicator can be transmitted to one or more upstream nodes to indicate congestion. As a further example, the congestion indicator can be a bit value such as a DE bit set to a value of "1," (e.g., DE=1).

In an aspect, when an upstream node receives the one or more frames with DE=1, the upstream node can transmit the received one or more frames with DE=1 to other upstream nodes without changing the DE bit value. If the upstream node that receives the one or more frames with DE=1 is a boundary node, the boundary node can reduce a data rate for downstream transmission and set DE bit to zero in reverse direction. In an aspect, the congestion indicator (e.g., DE bit) can be used to indicate congestion for port, connection, and/or class of service.

FIG. **1**B illustrates an exemplary packet or frame structure that can be implemented as part of the systems and methods of the present disclosure. In an aspect, the frame structure can be the same or similar to the standard frame structure defined by IEEE 802.3-2005, incorporated herein by reference in its entirety. As an example, the frame structure can comprise an Interframe Gap, P/SFD (Preamble/Start of Frame Delimiter), TCI (Tag Control Information=VID+CFI/DE+PCP), S-Tag TPID: Service Tag Identifier, C-Tag TPID: Customer Tag Identifier, LIT (Length/Type): Length of frame or data type, user data, FCS (Frame Check Sequence), or Ext: Extension field, or a combination thereof.

FIG. **1**C illustrates an exemplary customer tag (C-Tag) TCI format that can be implemented as part of the systems and methods of the present disclosure. In an aspect, the customer tag format can be the same or similar to the standard frame structure defined by IEEE 802.3-2005. As an example, the customer tag format can comprise priority code point (PCP), canonical format indicator (CFI), or VLAN identifier, or a combination thereof. Other data points can be used.

FIG. **1**D illustrates an exemplary service tag (S-Tag) TCI format that can be implemented as part of the systems and methods of the present disclosure. In an aspect, the service tag format can be the same or similar to the standard frame structure defined by IEEE 802.3-2005. As an example, the service tag format can comprise priority code point (PCP), drop/discard eligibility indicator (DEI) such as discard eligibility bit (DE), or VLAN identifier, or tag protocol identifier (TPID), or a combination thereof. Other data points can be used.

In an aspect, DE bits in the service tag format in FIG. **1**D can be used to identify frame colors at an External Network-Network Interface (ENNI), if Differentiated Services Code Point (DSCP) is not used to identify colors. As an example, colors can be used to identify the bandwidth profile conformance of a particular of a data packet, such as a service frame. As an example, green can indicate conformance with a committed rate of the bandwidth profile. As another example, yellow can indicate non-conformance with committed rate, but conformance with an excess rate of the bandwidth profile. As a further example, red can indicate non-conformance with a committed rate and an excess rate of the bandwidth profile.

In an aspect, priority code point (PCP) can be adequate to represent less than eight Class of Services (CoS) and two colors (e.g., Green and Yellow). Accordingly, the DE bits in a

service tag format of frame structure can be used to indicate congestion end-to-end or on a domain basis. For example, when a downstream node is congested, the downstream node can define the DE bit of the frame structure to "1" and can transmit the frame structure to one or more upstream devices to indicate that the downstream node has determined a congested state.

In an exemplary aspect, the methods and systems can be implemented on a computing system such as computing device **201** as illustrated in FIG. **2** and described below. By way of example, one or more of the user device **102** and the computing device **104** of FIG. **1** can be a computer as illustrated in FIG. **2**. Similarly, the methods and systems disclosed can utilize one or more computers to perform one or more functions in one or more locations. FIG. **2** is a block diagram illustrating an exemplary operating environment for performing the disclosed methods. One skilled in the art will appreciate that this is a functional description and that the respective functions can be performed in a system by software, hardware, or a combination of software and hardware. This exemplary operating environment is only an example of an operating environment and is not intended to suggest any limitation as to the scope of use or functionality of operating environment architecture. Neither should the operating environment be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in the exemplary operating environment.

The present methods and systems can be operational with numerous other general purpose or special purpose computing system environments or configurations. Examples of well known computing systems, environments, and/or configurations that can be suitable for use with the systems and methods comprise, but are not limited to, personal computers, server computers, laptop devices, and multiprocessor systems. Additional examples comprise set top boxes, programmable consumer electronics, network PCs, minicomputers, mainframe computers, distributed computing environments that comprise any of the above systems or devices, and the like.

The processing of the disclosed methods and systems can be performed by software components. The disclosed systems and methods can be described in the general context of computer-executable instructions, such as program modules, being executed by one or more computers or other devices. Generally, program modules comprise computer code, routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. The disclosed methods can also be practiced in grid-based and distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules can be located in both local and remote computer storage media including memory storage devices.

Further, one skilled in the art will appreciate that the systems and methods disclosed herein can be implemented via a general-purpose computing device in the form of a computing device **201**. The components of the computing device **201** can comprise, but are not limited to, one or more processors or processing units **203**, a system memory **212**, and a system bus **213** that couples various system components including the processor **203** to the system memory **212**. In the case of multiple processing units **203**, the system can utilize parallel computing.

The system bus **213** represents one or more of several possible types of bus structures, including a memory bus or memory controller, a peripheral bus, an accelerated graphics port, and a processor or local bus using any of a variety of bus architectures. By way of example, such architectures can comprise an Industry Standard Architecture (ISA) bus, a Micro Channel Architecture (MCA) bus, an Enhanced ISA (EISA) bus, a Video Electronics Standards Association (VESA) local bus, an Accelerated Graphics Port (AGP) bus, and a Peripheral Component Interconnects (PCI), a PCI-Express bus, a Personal Computer Memory Card Industry Association (PCMCIA), Universal Serial Bus (USB) and the like. The bus **213**, and all buses specified in this description can also be implemented over a wired or wireless network connection and each of the subsystems, including the processor **203**, a mass storage device **204**, an operating system **205**, network software **206**, network data **207**, a network adapter **208**, system memory **212**, an Input/Output Interface **210**, a display adapter **209**, a display device **211**, and a human machine interface **202**, can be contained within one or more remote computing devices **214***a,b,c* at physically separate locations, connected through buses of this form, in effect implementing a fully distributed system.

The computing device **201** typically comprises a variety of computer readable media. Exemplary readable media can be any available media that is accessible by the computing device **201** and comprises, for example and not meant to be limiting, both volatile and non-volatile media, removable and non-removable media. The system memory **212** comprises computer readable media in the form of volatile memory, such as random access memory (RAM), and/or non-volatile memory, such as read only memory (ROM). The system memory **212** typically contains data such as network data **207** and/or program modules such as operating system **205** and network software **206** that are immediately accessible to and/or are presently operated on by the processing unit **203**.

In another aspect, the computing device **201** can also comprise other removable/non-removable, volatile/non-volatile computer storage media. By way of example, FIG. **2** illustrates a mass storage device **204** which can provide non-volatile storage of computer code, computer readable instructions, data structures, program modules, and other data for the computing device **201**. For example and not meant to be limiting, a mass storage device **204** can be a hard disk, a removable magnetic disk, a removable optical disk, magnetic cassettes or other magnetic storage devices, flash memory cards, CD-ROM, digital versatile disks (DVD) or other optical storage, random access memories (RAM), read only memories (ROM), electrically erasable programmable read-only memory (EEPROM), and the like.

Optionally, any number of program modules can be stored on the mass storage device **204**, including by way of example, an operating system **205** and network software **206**. Each of the operating system **205** and network software **206** (or some combination thereof) can comprise elements of the programming and the network software **206**. Network data **207** can also be stored on the mass storage device **204**. Network data **207** can be stored in any of one or more databases known in the art. Examples of such databases comprise, DB2®, Microsoft® Access, Microsoft® SQL Server, Oracle®, mySQL, PostgreSQL, and the like. The databases can be centralized or distributed across multiple systems.

In another aspect, the user can enter commands and information into the computing device **201** via an input device (not shown). Examples of such input devices comprise, but are not limited to, a keyboard, pointing device, mouse, microphone, joystick, scanner, tactile input devices such as gloves, and other body coverings, and the like These and other input devices can be connected to the processing unit **203** via a human machine interface **202** that is coupled to the system

bus **213**, but can be connected by other interface and bus structures, such as a parallel port, game port, an IEEE 1394 Port (also known as a Firewire port), a serial port, or a universal serial bus (USB).

In yet another aspect, a display device **211** can also be connected to the system bus **213** via an interface, such as a display adapter **209**. It is contemplated that the computing device **201** can have more than one display adapter **209** and the computer **201** can have more than one display device **211**. For example, a display device can be a monitor, an LCD (Liquid Crystal Display), or a projector. In addition to the display device **211**, other output peripheral devices can comprise components such as speakers (not shown) and a printer (not shown) which can be connected to the computing device **201** via Input/Output Interface **210**. Any step and/or result of the methods can be output in any form to an output device. Such output can be any form of visual representation, including, but not limited to, textual, graphical, animation, audio, tactile, and the like. The display **211** and computing device **201** can be part of one device, or separate devices.

The computing device **201** can operate in a networked environment using logical connections to one or more remote computing devices **214***a,b,c*. By way of example, a remote computing device can be a personal computer, portable computer, a smart phone, a server, a router, a network computer, a peer device or other common network node, and so on. Logical connections between the computing device **201** and a remote computing device **214***a,b,c* can be made via a network **215**, such as a local area network (LAN) and a general wide area network (WAN). Such network connections can be through a network adapter **208**. A network adapter **208** can be implemented in both wired and wireless environments. Such networking environments are conventional and commonplace in dwellings, offices, enterprise-wide computer networks, intranets, and the Internet.

For purposes of illustration, application programs and other executable program components such as the operating system **205** are illustrated herein as discrete blocks, although it is recognized that such programs and components reside at various times in different storage components of the computing device **201**, and are executed by the data processor(s) of the computer. An implementation of software **206** can be stored on or transmitted across some form of computer readable media. Any of the disclosed methods can be performed by computer readable instructions embodied on computer readable media. Computer readable media can be any available media that can be accessed by a computer. By way of example and not meant to be limiting, computer readable media can comprise "computer storage media" and "communications media." "Computer storage media" comprise volatile and non-volatile, removable and non-removable media implemented in any methods or technology for storage of information such as computer readable instructions, data structures, program modules, or other data. Exemplary computer storage media comprises, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by a computer.

The methods and systems can employ Artificial Intelligence techniques such as machine learning and iterative learning. Examples of such techniques include, but are not limited to, expert systems, case based reasoning, Bayesian networks, behavior based AI, neural networks, fuzzy systems, evolutionary computation (e.g. genetic algorithms), swarm intelligence ant algorithms), and hybrid intelligent systems (e.g. expert inference rules generated through a neural network or production rules from statistical learning).

FIGS. **3-4** illustrate an exemplary system and network. In an aspect, plurality of nodes **302***a*, **302***b*, **302***c*, **302***d*, **302***e*, **302***f* can be in communication with one or more user devices **303***a*, **303***b* and/or one or more computing devices **304***a*, **304***b*. One or more of the nodes **302***a*, **302***b*, **302***c*, **302***d*, **302***e*, **302***f* can be a network access point, router, switch, or other communication device. As an example, one or more user devices **303** can be an electronic device such as a computer, a smartphone, a laptop, a tablet, a set top box, a display device, or other device capable of communicating with one or more of the nodes **302***a*, **302***b*, **302***c*, **302***d*, **302***e*, **302***f* of the network. As a further example, one or more computing devices **304***a*, **304***b* can be a server, a gateway, customer premises equipment (CPE), network interface device (NID), optical networking unit (ONU), headend, terminal server, cable modem terminal system, or other network device. As an example, one or more of the nodes **302***a*, **302***b*, **302***c*, **302***d*, **302***e*, **302***f* can be configured to communicate with another of the nodes **302***a*, **302***b*, **302***c*, **302***d*, **302***e*, **302***f* and/or one or more of the computing devices **304***a*, **304***b* via one or more communication paths **306**. In an aspect, the one or more communication paths **306** can comprise one or more uninterrupted communication links, sequential links, pre-defined paths or links, and/or intervening nodes. Links can comprise a single point-to-point connection between two devices or access points. Paths can comprise one or more links. As an example, one or more of the communication paths **306** can comprise one or more of the nodes **302***a*, **302***b*, **302***c*, **302***d*, **302***e*, **302***f*. As a further example, one or more of the nodes **302***a*, **302***b*, **302***c*, **302***d*, **302***e*, **302***f* can be configured as a mesh network. In an aspect, one or more of the communication paths **306** can be configured to transmit one or inure services.

In an aspect, one or more path elements **308***a*, **308***b*, **308***c*, **308***d*, **308***e*, **308***f* can comprise information relating to one or more of the communication paths **306**. One or more path elements **308***a*, **308***b*, **308***c*, **308***d*, **308***e*, **308***f* can comprise information relating to congestion, path priority, path cost, capacity, bandwidth, signal strength, latency, error rate, path usage, and the like. As an example, the path element **308***a*, **308***b*, **308***c*, **308***d*, **308***e*, **308***f* can be or comprise the congestion element **118** (FIG. **1**A). As a further example, the path element **308***a*, **308***b*, **308***c*, **308***d*, **308***e*, **308***f* can be configured to determine and/or generate a congestion indicator **310** such as a value of the DE bit in a received network frame.

As illustrated in FIG. **3**, congestion can occur downstream a boundary node such as node **302***b*. As an example, FIG. **3** illustrates congestion occurring at computing device **304***a*. In an aspect, congestion information can be determined on a per port basis for a given device (e.g., node). For example, a port of a boundary node can receive a frame with congestion information (e.g., DE=1 due to downstream congestion). Although the port of the boundary node is not congested, a reduction in data rate transmission from the end node port may be required to address network congestion. As an example, the port receiving the congestion information can be set to a congested state in order to trigger a congestion control algorithm. As a further example, the PAUSE flow control defined by IEEE 802.3, hereby incorporated herein by reference, standards can be used as a congestion control algorithm. In an aspect, congestion information can be determined based on various entities such as on a connection (e.g., Ethernet connection, Ethernet virtual connection, etc.) basis and/or class of service (CoS) flow basis.

As illustrated in FIG. **4**, congestion can occur at a boundary node such as boundary node **302***b*. In an aspect, congestion at node **302***b* can occur when a buffer associated with the node **302***b* exceed a threshold. As an example, a buffer associated with a port (e.g., access port or network port) of the node **302***b* can exceed a pre-set threshold (T_B). Accordingly, the port itself is congested, and the node **302***b* can activate a congestion control algorithm to reduce a data rate in both directions for all connections transported on that port. As a further example, the node **302***b* can then transmit frames having DE=0.

In an aspect, a buffer associated with a connection (e.g., Ethernet connection, virtual connection, etc.) of the node **302***b* can exceed a pre-set threshold (t_B). Accordingly, the connection itself is congested, and the node **302***b* can activate a congestion control algorithm to reduce a data rate in both directions for the connection transported on that port. As a further example, the node **302***b* can then transmit frames having DE=0.

In an aspect, a buffer associated with a class of service of the node **302***b* can exceed a pre-set threshold (t_{CB}). Accordingly, the class of service itself is congested, and the node **302***b* can activate a congestion control algorithm to reduce a data rate in one or both directions for that specific CoS flow transported on that port. As a further example, the node **302***b* can then transmit frames having DE=0.

In an aspect, when congestion indication is received at a node, the transmission data rate can be modified in response to the congestion information. As an example, the transmission data rate can be configured based upon one or more of the following formulas:

$$BWEF^{ASF} = \sum_{n=1}^{N} BWEF_n^{SF}; \quad \text{(EQ 1)}$$

if it is port level congestion,

where $BWEF^{SF}{}_n$ represents the effective bandwidth of a service flow and $BWEF^{ASF}$ represents effective bandwidth for the aggregated service flow of one or more service flows.

In an aspect, if congestion persists after the data rate is configured by EQ1 or the current rate is already at the effective bandwidth, then the congested node can reduce its transmission data rate to

$$CIR^{ASF} = \sum_{n=1}^{N} CIR_n^{SF}; \quad \text{(EQ 2)}$$

if it is port level congestion,

where $CIR^{SF}{}_n$ represents the committed information rate of a service flow, and CIR^{ASF} represents the committed information rate for the aggregate service flow of one or more service flows.

As an example, one or more nodes can discard (e.g. delete, do not transmit, etc.) yellow frames first and then green frames to reduce frames in the buffer below threshold T_B for port, or t_B for connection or t_{CB} for CoS flow. As a further example, thresholds (T_B, t_B, t_{CB}) can be defined such that a buffer size that is determined to be below the defined thresholds can trigger a reduction in transmission data rates to an effective bandwidth rate or to a committed information rate.

In an aspect, transmitted and received frame counts can be monitored and correction can be made to ensure the updated transmission data rates. As an example, in order to allow statistical multiplexing with agreed packet loss ratio or frame loss ratio (FLR), effective bandwidth can be used in allocating bandwidth. As a further example, effective bandwidth can be calculated for each service flow (SF) and a sum of the effective bandwidths can be assigned to aggregate service flow (ASF). In an aspect, a service flow can comprise an end-to-end traffic flow defined by traffic parameters such as average output rate, maximum output burst, and the like. As an example, a service flow can comprise an Ethernet virtual connection between user network interfaces (UNI) of a network. As a further example, a service flow can comprise a group of packets/frames flowing in an Ethernet virtual connection between UNI's and belong to an application with a defined class of service. An aggregate service flow can comprise one or more service flows.

In an aspect, effective bandwidth can be calculated based on one or more of the following formulas:

$$BWEF^{ASF} = \sum_{n=1}^{N} BWEF_n^{SF}; \quad \text{(EQ 3)}$$

instead of

$$\text{Total Bandwidth} = CIR^{ASF} + EIR^{ASF}, \quad \text{(EQ 4)}$$

where

$$CIR^{ASF} = \sum_{n=1}^{N} CIR_n^{SF}; \quad \text{(EQ 5)}$$

$$EIR^{ASF} = \sum_{n=1}^{N} EIR_n^{SF}; \quad \text{(EQ 6)}$$

where CIR can be defined as average output rate of the shaper or a policer for green frames, CBS can be defined as maximum output burst of the shaper or a policer for green frames, EIR can be defined as average output rate of the shaper or policer for yellow frames, and EBS can be defined as maximum output burst of the shaper or policer for yellow frames.

In an aspect, Effective Bandwidth for a given SF can be calculated as

$$BWEF^{SF}=\max(CIR,PR/(1+(\max(\text{jitter}_{A\text{-}Z},\text{jitter}_{Z\text{-}A})*PR)/MBS), \quad \text{(EQ7).}$$

where PR=CIR+EIR

Equation EQ7 can be valid for service flows (or EVCs or CoS) with no loss SLAs (service level agreements). In an aspect, Effective Bandwidth for a given service flow (or EVCs or CoS) with loss SLAs, can be calculated as:

$$BWEF^{SF}=\max(CIR, PR*\alpha) \quad \text{(EQ8),}$$

wherein $\alpha=(\beta-b)+\sqrt{(\beta-b)\hat{}2+4(CIR/PR)\beta*b/2\beta}$,

wherein $\beta=(1n1/FLR)(MBS/CIR)(1-CIR/PR)PR$, and

wherein $Jitter_{A\text{-}Z}$ is a delay variation between a pair of packets of a given connection or flow travelling from end point A of the connection or flow to the end point Z of the connection or flow. $Jitter_{Z\text{-}A}$, is the delay variation between a pair of packets of a given connection or flow travelling from the end point Z of the connection or flow to the end point A of the connection or flow. PR is the peak rate and can be defined by a port rate or EIR+CIR for a given connection or flow. FLR is the frame loss ratio and can be defined by ((the number of

frames transmitted minus the number of frames received)/(number of frames transmitted)) in a given connection or flow. MBS is the maximum output burst of a shaper or policer associated with the PR.

In an aspect, provided are methods for providing services to a user and/or user device. FIG. **5** illustrates an exemplary method for managing a network. In step **502**, congestion information, e.g., a frame with a service tag can be received at a first node. In an aspect, the service tag can represent congestion information of at least a portion of the network. As an example, the service tag can comprise discard eligibility data representing the congestion information. In an aspect, the discard eligibility data can represent the congestion information. As an example, the discard eligibility bit having a value of "1" can indicate a congested state of at least a portion of the network. As a further example, the discard eligibility bit having a value of "0" can indicate a non-congested state of at a portion of the network. In an aspect, the congestion information can represent one or more of port level capacity, connection level capacity, and class of service level capacity.

In step **503**, a determination can be made if a node (e.g., the first node) is a boundary node.

In step **504**, if the first node is a boundary node, a downstream data rate can be modified based upon the congestion information. In an aspect, modifying a downstream data rate can comprise reducing a downstream data rate based upon a congestion control algorithm. As an example, modifying a downstream data rate can comprise configuring the transmission data rate of one or more nodes based upon the formulas set forth herein.

In step **506**, if the first node is not a boundary node, the congestion information can be transmitted to a second node. As an example, the second node is upstream of the first node. In an aspect, the congestion information can be received at the second node. If the second node is a boundary node, a downstream data rate can be modified based upon the congestion information. If the second node is not a boundary node, transmitting the congestion information to a third node and the process can be repeated.

In an aspect, FIG. **6** illustrates an exemplary method for determining congestion information at a node. In step **602**, congestion information can be determined for at least a portion of a network. In an aspect, determining congestion information can comprise receiving an upstream communication from a downstream device. As an example, determining congestion information can comprise comparing a network parameter to a threshold value. As a further example, the network parameter is buffer capacity, effective bandwidth, port level capacity, connection level capacity, or class of service level capacity, or a combination thereof.

In step **604**, a service tag in one or more packets or frames can be set to represent the congestion information. In an aspect, the service tag can comprise discard eligibility data representing the congestion information. As an example, the discard eligibility data representing the congestion information can be a binary bit. As another example, the discard eligibility bit having a value of one can indicate a congested state of the at least a portion of the network. As a further example, the discard eligibility bit having a value of zero can indicate a non-congested state of the at least a portion of the network.

In step **606**, the one or more frames with the service tag can be transmitted to an upstream device of the network. In an aspect, the upstream device can receive the frames with service tag. As an example, if the upstream device is a boundary node, a downstream data rate can be modified based upon the congestion information. As a further example, if the upstream device is not a boundary node, the congestion information can be transmitted to a second upstream device.

In an aspect, FIG. **7** illustrates an exemplary method for determining or modifying an effective bandwidth and/or data rate. In step **702**, an indication of network congestion can be received. In an aspect, the indication of network congestion can relate to a service flow. As an example, a service tag can be provided as the indication of network congestion. In an aspect, the service tag can comprise discard eligibility data representing the congestion information. As an example, the discard eligibility data representing the congestion information can be a binary bit. As another example, the discard eligibility bit having a value of one can indicate a congested state of the at least a portion of the network. As a further example, the discard eligibility bit having a value of zero can indicate a non-congested state of the at least a portion of the network. In an aspect, the indication of network congestion can represent one or more of port level capacity, connection level capacity, and class of service level capacity.

In step **704**, an effective bandwidth can be determined for the service flow. In an aspect, the effective bandwidth can be determined based upon the formulas disclosed herein.

In step **706**, a data rate associated with the service flow can be modified based upon the effective bandwidth determined. In an aspect, modifying the data rate can comprise reducing a data rate based upon a congestion control algorithm. As an example, modifying the data rate can comprise configuring the transmission data rate of one or more nodes based upon the formulas set forth herein.

The systems and methods of the present disclosure can maximize network utilization by regulating traffic between source and destination automatically, thereby reducing substantial delays and frame drops.

While the methods and systems have been described in connection with preferred embodiments and specific examples, it is not intended that the scope be limited to the particular embodiments set forth, as the embodiments herein are intended in all respects to be illustrative rather than restrictive.

Unless otherwise expressly stated, it is in no way intended that any method set forth herein be construed as requiring that its steps be performed in a specific order. Accordingly, where a method claim does not actually recite an order to be followed by its steps or it is not otherwise specifically stated in the claims or descriptions that the steps are to be limited to a specific order, it is no way intended that an order be inferred, in any respect. This holds for any possible non-express basis for interpretation, including: matters of logic with respect to arrangement of steps or operational flow; plain meaning derived from grammatical organization or punctuation; the number or type of embodiments described in the specification.

It will be apparent to those skilled in the art that various modifications and variations can be made without departing from the scope or spirit. Other embodiments will be apparent to those skilled in the art from consideration of the specification and practice disclosed herein. It is intended that the specification and examples be considered as exemplary only, with a true scope and spirit being indicated by the following claims.

What is claimed is:

1. A method comprising:

receiving a message at a first node, wherein the message comprises an upstream communication comprising one or more frames from a downstream device, and at least one of the one or more frames in the message comprises a service tag,

wherein the service tag represents congestion information of at least a portion of a network, and

wherein the message has traversed the portion of the network;

if the first node is a boundary node, modifying a downstream data rate based upon the congestion information; and

if the first node is not a boundary node, transmitting the congestion information to a second node, wherein the second node is upstream of the first node.

2. The method of claim **1**, wherein the service tag comprises discard eligibility data representing the congestion information.

3. The method of claim **2**, wherein the discard eligibility data representing the congestion information comprises a binary bit.

4. The method of claim **3**, wherein the binary bit having a value of one indicates a congested state of the at least a portion of the network.

5. The method of claim **3**, wherein the binary bit having a value of zero indicates a non-congested state of the at least a portion of the network.

6. The method of claim **1**, wherein the congestion information represents port level capacity.

7. The method of claim **1**, wherein modifying a downstream data rate comprises reducing a downstream data rate based upon a congestion control algorithm.

8. The method of claim **1**, wherein the second node is upstream of the first node.

9. The method of claim **1**, further comprising receiving the congestion information at the second node; if the second node is a boundary node, modifying a downstream data rate based upon the congestion information; and if the second node is not a boundary node, transmitting the congestion information to a third node, wherein the third node is upstream of the second node.

10. A method comprising:

determining congestion information of at least a portion of a network;

receiving a message, wherein the message comprises an upstream communication comprising one or more frames from a downstream device;

setting a service tag in at least one of the one or more frames in the message to represent the congestion information; and

transmitting the service tag to a first upstream device of the network in the message.

11. The method of claim **10**, wherein determining the congestion information comprises comparing a network parameter to a threshold value.

12. The method of claim **11**, wherein the network parameter comprises effective bandwidth.

13. The method of claim **10**, wherein the service tag comprises discard eligibility data representing the congestion information.

14. The method of claim **13**, wherein the discard eligibility data representing the congestion information comprises a binary bit.

15. The method of claim **14**, wherein the binary bit having a value of one indicates a congested state of the at least a portion of the network.

16. The method of claim **14**, wherein the binary bit having a value of zero indicates a non-congested state of the at least a portion of the network.

17. The method of claim **10**, further comprising receiving the congestion information at the first upstream device; if the first upstream device is a boundary node, modifying a downstream data rate based upon the congestion information; and if the first upstream device is not a boundary node, transmitting the congestion information to a second upstream device.

18. A method comprising:

receiving a first indication of network congestion in a service flow via a message in the service flow, wherein the message comprises a communication comprising one or more frames comprising a service tag that represents the service flow and the network congestion relating to the service flow;

determining an effective bandwidth for the service flow; and

modifying a data rate associated with the service flow based upon the effective bandwidth.

19. The method of claim **18**, wherein the service tag comprises discard eligibility data representing the indication of network congestion.

20. The method of claim **19**, wherein the discard eligibility data comprises a binary bit.

21. The method of claim **20**, wherein the binary bit having a value of one indicates a congested state of at least a portion of a network.

22. The method of claim **20**, wherein the binary bit having a value of zero indicates a non-congested state of at least a portion of a network.

23. The method of claim **18**, wherein the indication of network congestion represents connection level capacity.

24. The method of claim **18**, further comprising receiving a second indication of network congestion, the second indication relating to the service flow; determining a committed information rate for the service flow; and modifying the data rate associated with the service flow based upon the committed information rate.

25. The method of claim **1**, wherein the boundary node comprises a node that initiates congestion control.

26. The method of claim **1**, wherein the boundary node comprises a node that terminates congestion control.

27. The method of claim **1**, wherein the boundary node comprises a node that supports Ethernet virtual connection termination points.

28. The method of claim **1**, wherein the boundary node comprises a provider edge router.

29. The method of claim **1**, wherein the congestion information represents connection level capacity.

30. The method of claim **1**, wherein the congestion information represents class of service level capacity.

31. The method of claim **17**, wherein the boundary node comprises a node that supports Ethernet virtual connection termination points.

32. The method of claim **17**, wherein the boundary node comprises a provider edge router.

* * * * *